

## METHODS AND NUCLEIC ACIDS FOR THE ANALYSIS OF COLON CELL PROLIFERATIVE DISORDERS

### CROSS-REFERENCE TO RELATED APPLICATIONS

5 This application claims the benefit of priority to United States Patent Application Serial No. 10/603,138, filed 23 June, 2003 and entitled METHODS AND NUCLEIC ACIDS FOR ANALYSES OF COLORECTAL CELL PROLIFERATIVE DISORDERS, which is incorporated herein by reference in its entirety.

### 10 FIELD OF THE INVENTION

Aspects of the present invention relate to cancer, and to the detection and progression of cancer. More particular aspects relate to nucleic acids and kits having diagnostic, prognostic and therapeutic utility for detecting and distinguishing colon cell proliferative disorders, based on methylation patterns of relevant genomic DNA sequences.

### 15 BACKGROUND

In the United States the annual incidence of colorectal cancer is approximately 150,000, with 56,600 individuals dying from colorectal cancer each year. The lifetime risk of colorectal cancer in the general population is about 5 to 6 percent. Despite intensive efforts in recent years  
20 in screening and early detection of colon cancer, until today most cases are diagnosed in an advanced stage with regional or distant metastasis. While the therapeutic options include surgery and adjuvant or palliative chemotherapy, most patients die from progression of their cancer within a few months. Identifying the molecular changes that underly the progression of colon cancer and the formation of metastasis may help to develop new diagnostic and  
25 therapeutic options that could improve the overall poor prognosis of these patients.

The current model of colorectal pathogenesis favours a stepwise progression of adenomas which includes the development of dysplasia and finally signs of invasive cancer. The molecular changes underlying this adenoma-carcinoma sequence include genetic and epigenetic alterations of tumor suppressor genes (APC, p53, DCC), the activation of oncogenes  
30 (K-ras) and the inactivation of DNA mismatch repair genes<sup>1</sup>. Recently, further molecular changes and genetic defects have been revealed. Thus, activation of the Wnt signalling pathway not only includes mutations of the APC gene, but may also result from beta-catenin mutations<sup>5</sup>. Furthermore, alterations in the TGF-beta signalling pathway together with its signal transducers SMAD4 and SMAD2 have been linked to the development of colon cancer.

35 Despite recent progress in the understanding of the pathogenesis of adenomas and

carcinomas of the colon and their genetic and molecular changes, the genetic and epigenetic changes underlying the development of metastasis are less well understood. It is, however, generally well accepted that the process of invasion and proteolysis of the extracellular matrix, as well as infiltration of the vascular basement membrane involve adhesive proteins, such as members of the family of integrin receptors, the cadherins, the immunoglobulin superfamily, the laminin binding protein and the CD44 receptor. Apart from adhesion, the process of metastasis formation also includes the induction and regulation of angiogenesis (VEGF, bFGF), the induction of cell proliferation (EGF, HGF, IGF) and the activation of proteolytic enzymes (MMPs, TIMPs, uPAR), as well as the inhibition of apoptosis (Bcl-2, Bcl-X). More recently other groups have compared the genetic and molecular changes in metastatic lesions to the changes found in primary colorectal cancers. Thus, Kleeff et al. reported the loss of DOC-2, a candidate tumor suppressor gene, both in primary and metastatic colorectal cancer. Furthermore, Zauber et al. reported that in their series of 42 colorectal cancers Ki-ras mutations in the primary cancers were identical in all of the 42 paired primary and synchronous metastatic lesions. Similarly loss of heterozygosity at the APC locus was identical for 39 paired carcinomas and synchronous metastasis. The authors concluded that for Ki-ras and APC genes the genetic changes in metastasis are identical to the primary colorectal cancer. However, other groups have found genetic and molecular changes in metastatic colon cancers, that are not present in the primary cancers. Thus, the development of LOH of chromosome 3p in colorectal metastasis has been reported. In addition, using comparative genomic hybridization several alterations were found in liver metastasis that were unique to metastatic lesions (-9q, -11q, and -17q)<sup>38</sup>.

Apart from mutations aberrant methylation of CpG islands has been shown to lead to the transcriptional silencing of certain genes that have been previously linked to the pathogenesis of various cancers. CpG islands are short sequences which are rich in CpG dinucleotides and can usually be found in the 5' region of approximately 50% of all human genes. Methylation of the cytosines in these islands leads to the loss of gene expression and has been reported in the inactivation of the X chromosome and genomic imprinting. Recently several groups have also analysed the methylation of various genes in colorectal cancer and reported the transcriptional silencing by promoter methylation for p16INK4, p14ARF, p15INK4b, MGMT, hMLH1, GSTP1, DAPK, CDH1, TIMP-3 and APC among others. Thus apart from mutational inactivation of certain genes, the hypermethylation of these genes also contributes significantly to the pathogenesis of this disease.

In recent years several genes that are methylated in colon cancer have been identified by MS AP-PCR. This group of genes among others, includes TPEF/HPP1 which is frequently

5 methylated in colon cancers and which was independently identified by two different groups using the MS AP-PCR method. See for example, Young J, Biden KG, Simms LA, Huggard P, Karamatic R, Eyre HJ, Sutherland GR, Herath N, Barker M, Anderson GJ, Fitzpatrick DR, Ramn GA, Jass JR, Leggett BA. HPP1: a transmembrane protein-encoding gene commonly methylated in colorectal polyps and cancers. Proc Natl Acad Sci USA 2001;98:265-270.

10 ALX4 gene methylation was identified using differential methylation hybridization in a study by Yan et al. of genome-wide screening for CpG island hypermethylation in breast cancer samples. In their study ALX4 gene methylation was most prominent in poorly differentiated breast cancers (Yan PS, Perry MR, Laux DE, Asare AL, Caldwell CW, Huang TH. CpG island arrays: an application toward deciphering epigenetic signatures of breast cancer. Clin Cancer Res 2000;6:1432-1438.). ALX4 is a putative transcription factor that belongs to the family of paired-class homeoproteins. This gene is part of a family of genes that includes the mammalian genes Alx3, Cart-1, MHox, and S8 and exhibits similarity to the *Drosophila* gene *aristaless*. It binds palindromic DNA sequences (5'-TAAT-3') as either homodimers or as heterodimers with  
15 other family members and strongly activates transcription from a promoter containing the homeodomain binding site, P2. ALX4 is expressed at several sites during development, including the craniofacial and limb-bud mesenchyme. Interestingly, ALX4 deficient mice exhibit body-wall defects, preaxial polydactyly, and a decreased size of the parietal plate of the skull, while mutations of the human homeobox gene ALX4 have been found in inherited defects of  
20 skull ossification. ALX4 is also expressed in various tissues whose development is dependent on epithelial-mesenchymal interactions and regulates mesenchymal-specific activities of LEF-1.

*Multifactorial approach.* Cancer diagnostics has traditionally relied upon the detection of single molecular markers (e.g. gene mutations, elevated PSA levels). Unfortunately, cancer is a disease state in which single markers have typically failed to detect or differentiate many forms  
25 of the disease. Thus, assays that recognize only a single marker have been shown to be of limited predictive value. A fundamental aspect of this invention is that methylation based cancer diagnostics and the screening, diagnosis, and therapeutic monitoring of such diseases will provide significant improvements over the state-of-the-art that uses single marker analyses by the use of a selection of multiple markers. The multiplexed analytical approach is particularly  
30 well suited for cancer diagnostics since cancer is not a simple disease, this multi-factorial "panel" approach is consistent with the heterogeneous nature of cancer, both cytologically and clinically.

Key to the successful implementation of a panel approach to methylation based diagnostic tests is the design and development of optimized panels of markers that can

characterize and distinguish disease states. This patent application describes an efficient and unique panel of genes the methylation analysis of one or a combination of the members of the panel enabling the detection of cell proliferative disorders of the prostate with a particularly high sensitivity, specificity and/or predictive value.

5        *Development of medical tests.* Two key evaluative measures of any medical screening or diagnostic test are its sensitivity and specificity, which measure how well the test performs to accurately detect all affected individuals without exception, and without falsely including individuals who do not have the target disease (predictive value). Historically, many diagnostic tests have been criticized due to poor sensitivity and specificity.

10        A true positive (TP) result is where the test is positive and the condition is present. A false positive (FP) result is where the test is positive but the condition is not present. A true negative (TN) result is where the test is negative and the condition is not present. A false negative (FN) result is where the test is negative but the condition is present. In this context: Sensitivity =  $TP/(TP+FN)$ ; Specificity =  $TN/(FP+TN)$ ; and Predictive value =  $TP/(TP+FP)$ .

15        Sensitivity is a measure of a test's ability to correctly detect the target disease in an individual being tested. A test having poor sensitivity produces a high rate of false negatives, *i.e.*, individuals who have the disease but are falsely identified as being free of that particular disease. The potential danger of a false negative is that the diseased individual will remain undiagnosed and untreated for some period of time, during which the disease may progress to a  
20        later stage wherein treatments, if any, may be less effective. An example of a test that has low sensitivity is a protein-based blood test for HIV. This type of test exhibits poor sensitivity because it fails to detect the presence of the virus until the disease is well established and the virus has invaded the bloodstream in substantial numbers. In contrast, an example of a test that has high sensitivity is viral-load detection using the polymerase chain reaction (PCR). High  
25        sensitivity is achieved because this type of test can detect very small quantities of the virus. High sensitivity is particularly important when the consequences of missing a diagnosis are high.

      Specificity, on the other hand, is a measure of a test's ability to identify accurately patients who are free of the disease state. A test having poor specificity produces a high rate of  
30        false positives, *i.e.*, individuals who are falsely identified as having the disease. A drawback of false positives is that they force patients to undergo unnecessary medical procedures treatments with their attendant risks, emotional and financial stresses, and which could have adverse effects on the patient's health. A feature of diseases which makes it difficult to develop diagnostic tests with high specificity is that disease mechanisms, particularly in cancer, often involve a plurality



of genes and proteins. Additionally, certain proteins may be elevated for reasons unrelated to a disease state. An example of a test that has high specificity is a gene-based test that can detect a p53 mutation. Specificity is important when the cost or risk associated with further diagnostic procedures or further medical intervention are very high.

5        *Bisulfite modification of DNA is an art-recognized tool used to assess CpG methylation status.* 5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine  
10 positions cannot be identified by sequencing, because 5-methylcytosine has the same base pairing behavior as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during, *e.g.*, PCR amplification.

The most frequently used method for analyzing DNA for the presence of 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine whereby, upon  
15 subsequent alkaline hydrolysis, cytosine is converted to uracil which corresponds to thymine in its base pairing behavior. Significantly, however, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is *converted* in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridization behavior, can now be detected as the only remaining cytosine using standard, art-recognized  
20 molecular biological techniques, for example, by amplification and hybridization, or by sequencing. All of these techniques are based on differential base pairing properties, which can now be fully exploited.

The prior art, in terms of sensitivity, is defined by a method comprising enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of  
25 the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of art-recognized methods for detecting 5-methylcytosine is  
30 provided by Rein, T., et al., *Nucleic Acids Res.*, 26:2255, 1998.

The bisulfite technique, barring few exceptions (*e.g.*, Zeschnigk M, et al., *Eur J Hum Genet.* 5:94-98, 1997), is currently only used in research. In all instances, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment, and either completely sequenced (Olek & Walter, *Nat Genet.* 1997 17:275-6, 1997), subjected to one or

more primer extension reactions (Gonzalzo & Jones, *Nucleic Acids Res.*, 25:2529-31, 1997; WO 95/00669; U.S. Patent No. 6,251,594) to analyze individual cytosine positions, or treated by enzymatic digestion (Xiong & Laird, *Nucleic Acids Res.*, 25:2532-4, 1997). Detection by hybridization has also been described in the art (Olek et al., WO 99/28498). Additionally, use of  
5 the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg & Clark, *Bioessays*, 16:431-6, 1994; Zeschnigk M, et al., *Hum Mol Genet.*, 6:387-95, 1997; Feil R, et al., *Nucleic Acids Res.*, 22:695-, 1994; Martin V, et al., *Gene*, 157:261-4, 1995; WO 9746705 and WO 9515373).

*Bisulfite Methylation Assay Procedures.* Various methylation assay procedures are  
10 known in the art, and can be used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (e.g., CpG islands) within a DNA sequence. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction enzymes.

15 For example, genomic sequencing has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, e.g., the method described by Sadri & Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction  
20 Analysis) (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

*COBRA.* COBRA<sup>TM</sup> analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated  
25 DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization  
30 probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples.

Other assays used in the art include "MethyLight<sup>TM</sup>" (a fluorescence-based real-time

PCR technique) (Eads et al., *Cancer Res.* 59:2302-2306, 1999), Ms-SNuPE<sup>TM</sup> (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalzo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146), and methylated CpG island amplification  
5 ("MCA"; Toyota et al., *Cancer Res.* 59:2307-12, 1999). These may be used alone or in combination with other of these methods.

*MethyLight.* The MethyLight<sup>TM</sup> assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqMan®) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 59:2302-2306, 1999).  
10 Briefly, the MethyLight<sup>TM</sup> process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an "unbiased" (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a "biased"  
15 (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

The MethyLight<sup>TM</sup> assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe  
20 hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlap any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control  
25 oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites.

The MethyLight<sup>TM</sup> process can be used with a "TaqMan®" probe in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes; e.g., with either biased  
30 primers and TaqMan® probe, or unbiased primers and TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq

polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Alternatively the MethylLight™ process can be used with 'Lightcycler' probes. A LightCycler™ probe is a pair of single-stranded fluorescent-labeled oligonucleotides. The first oligonucleotide probe is labeled at its 3'end with a donor fluorophore dye and the second is labeled at its 5'end with an acceptor fluorophore dyes. The free 3' hydroxyl group of the second probe is blocked with a phosphate group to prevent polymerase mediated extension.

During the annealing step of real-time quantitative PCR, the PCR primers and the LightCycler™ probes hybridize to their specific target regions causing the donor dye to come into close proximity to the acceptor dye. When the donor dye is excited by light, energy is transferred by Fluorescence Resonance Energy Transfer (FRET) from the donor to the acceptor dye. The energy transfer causes the acceptor dye to emit fluorescence wherein the increase of measured fluorescence signal is directly proportional to the amount of target DNA.

Typical reagents (e.g., as might be found in a typical MethylLight™-based kit) for MethylLight™ analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® and/or LightCycler™ probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

*Ms-SNuPE.* The Ms-SNuPE™ technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalzo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site(s) of interest. Small amounts of DNA can be analyzed (e.g., microdissected pathology sections), and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

Typical reagents (e.g., as might be found in a typical Ms-SNuPE™-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE™ primers for specific gene; reaction buffer (for the Ms-SNuPE™ reaction); and radioactive nucleotides. Additionally,



bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

*MSP.* MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers specific for methylated versus unmethylated DNA. This technique has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (e.g., as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

*Pronounced need in the art.* Therefore, in view of the incidence of colon cancer there is a substantial need in the art for the development of molecular markers that could be used for the early detection of colorectal cell proliferative disorders, in particular colon cancer. Additionally, there is a pronounced need in the art for the development of molecular markers that could be used to provide sensitive, accurate and non-invasive methods (as opposed to, e.g., biopsy) for the diagnosis, prognosis and treatment of colon cell proliferative disorders.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the results of the MS-APPCR analysis of pooled DNA from normal and adenoma tissues. A hypermethylated gene (ALX4) was detected in the adenoma DNA (box, arrow). R, *RsaI*; H, *RsaI/HpaII*; M, *RsaI/MspI*.

Figures 2A and 2B show the degree of methylation of the ALX4 gene as assessed by MethyLight™ assay as outlined in examples. The X-axis shows the percentage methylated reference (PMR). Figure A shows ALX4 gene methylation in normal colon mucosa and matched colon cancer; Figure B shows ALX4 gene methylation in primary colorectal cancers and

metastasis.

Figure 3 shows bisulfite sequencing. Four cases of cancers with a high degree of methylation and their matched normal non-malignant colon mucosa tissues were selected (inset). ALX4 gene methylation was confirmed by sequencing of bisulfite treated genomic DNA of samples 1-4. N, normal mucosa; T, tumor, numbers 1-4 correspond to cases 1-4 in the MethyLight assay (inset); MethyLight assay, indicates the CpG sites that were covered by both the MethyLight assay and sequencing of the respective DNA fragments.

Figures 4A and 4B show the number of methylated genes per primary colorectal cancer (A) and metastasis (B). CIMP+ status was defined as 2 or more methylated genes per patient.

Figures 5A, 5B and 5C show an analysis of methylation with regard to clinicopathological features of colorectal cancers. A. In cancers of the colon and the sigma methylation was significantly more frequent than in cancers of the rectum only. y axis: number of patients; x axis: number of methylated genes B. TPEF methylation was significantly more frequent in cancers of the colon and sigma. Number of patients analysed: colon cancer: n=23, sigma cancer: n=13, rectal cancer: n=10. One patient with recurrent cancer was not included in this analysis. y axis: number of patients; x axis: number of methylated genes. C. Cancers without distant metastasis were frequently methylated, grey box: no methylation; black box: methylation of 1 to 7 genes.

#### DETAILED DESCRIPTION OF THE INVENTION

For the purposes of the following invention the 'sensitivity' and 'specificity' refer to values calculated by reference to a sample set according to that described in the EXAMPLES disclosed herein.

##### Definitions:

The term "Observed/Expected Ratio" ("O/E Ratio") refers to the frequency of CpG dinucleotides within a particular DNA sequence, and corresponds to the [number of CpG sites / (number of C bases x number of G bases)] x band length for each fragment.

The term "CpG island" refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an "Observed/Expected Ratio" >0.6, and (2) having a "GC Content" >0.5. CpG islands are typically, but not always, between about 0.2 to about 1 kb in length.

The term "methylation state" or "methylation status" refers to the presence or absence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence. Methylation states at one or more particular palindromic CpG methylation sites (each having two CpG CpG dinucleotide sequences) within a DNA sequence include "unmethylated," "fully-

5 methylated" and "hemi-methylated."

The term "hemi-methylation" or "hemimethylation" refers to the methylation state of a palindromic CpG methylation site, where only a single cytosine in one of the two CpG dinucleotide sequences of the palindromic CpG methylation site is methylated (*e.g.*, 5'-NC<sup>M</sup>GN-3' (top strand); 3'-NGCN-5' (bottom strand)).

The term "hypermethylation" refers to the average methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

10 The term "hypomethylation" refers to the average methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

The term "microarray" refers broadly to both "DNA microarrays," and "DNA chip(s)," as recognized in the art, encompasses all art-recognized solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

"Epigenetic parameters" are, in particular, cytosine methylations. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analyzed using the described method but which, in turn, correlate with the DNA methylation.

25 The term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

The term "Methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

30 The term "MS AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

The term "MethyLight®" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

The term "HeavyMethyl<sup>TM</sup>" assay, in the embodiment thereof implemented herein, refers to a HeavyMethyl MethylLight® assay, which is a variation of the MethylLight® assay, wherein the MethylLight® assay is combined with methylation specific *blocking* probes covering CpG positions between the amplification primers.

5       The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

10       The term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

      The term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

15       The term "MCA" (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al., *Cancer Res.* 59:2307-12, 1999, and in WO 00/26401A1.

      The term "hybridization" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

20       "Stringent hybridization conditions," as defined herein, involve hybridizing at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognized equivalent thereof (e.g., conditions in which a hybridization is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognized equivalent  
25       thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, *Current Protocols in Molecular Biology*, (John Wiley & Sons, N.Y.) at Unit 2.10.

30       The term 'primary' when used in reference to cancer or other cell proliferative disorder shall be taken to mean the first to develop.

      The term 'metastasis' as used herein shall be taken to mean the transfer of a disease-producing agent (such as bacteria, cancer or other cell proliferative disorder cells) from an original site of disease to another part of the body with development of a similar lesion in the



new location.

Overview:

Despite intensive efforts to improve screening and early detection of colon cell  
5 proliferative disorders, most cases are diagnosed in an advanced stage with regional or distant  
metastasis which are associated with poor survival. The herein described invention discloses  
epigenetic markers that have novel utility for the analysis of colon cell proliferative disorders,  
combined with sensitive assay methods for the improved detection of said disorders. The  
invention presents improvements over the state of the art in that it provides a means for the  
10 detection of colon cell proliferative disorders by analysis of a gene panel, with a high sensitivity  
and specificity. The invention presents further improvements in that the 'gene panel' consists of  
at least one of seven genes and/or their regulatory sequences, thereby enabling a highly sensitive  
and specific but time and cost effective analysis. The invention further discloses particularly  
preferred combinations of said seven genes.

15 In one aspect, the present invention provides for the improved detection of colorectal  
carcinomas by determination of the methylation status of CpG dinucleotide sequences of the gene  
ALX4 and/or its regulatory sequences. In a further aspect the invention provides a further  
preferred means for the detection of colorectal carcinomas by determination of the methylation  
status of CpG dinucleotide sequences of the gene ALX4 and at least one gene selected from the  
20 group consisting TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their  
regulatory sequences.

In one aspect, the present invention provides for the use of the bisulfite technique, in  
combination with one or more methylation assays, for determination of the methylation status of  
CpG dinucleotide sequences of at least two genes taken from the group consisting ALX4, TPEF,  
25 p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences. It is  
particularly preferred that one of the genes is ALX4. According to the present invention,  
determination of the methylation status of CpG dinucleotide sequences within at least two  
members of said group of genes has diagnostic and prognostic utility. It is a further aspect of the  
invention that the analysed genomic sequences of the group consisting the group consisting  
30 ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory  
sequences are selected from the sequence listing (see Table 2).

In a further aspect the present invention provides a selection of genes consisting ALX4,  
TPEF and p16. Two or more of these genes are analysed in the form of a 'gene panel'. It is  
particularly preferred that one of the genes is ALX4. It is a further aspect of the invention that

the analysed genomic sequences are selected from the group consisting of ALX4, TPEF and p16 and/or their regulatory sequences are selected from the sequence listing (see Table 2).

Particular embodiments of the present invention provide a novel application of the analysis of methylation levels and/or patterns within said sequences that enables a precise  
5 detection, classification, treatment and overall prognosis of colon cell proliferative disorders. Early detection of colon cell proliferative disorders is directly linked with disease prognosis, and the disclosed method thereby enables the physician and patient to make better and more informed treatment decisions.

10 More particularly:

The present invention provides improved means for the detection of colorectal cell proliferative disorders. This aim is achieved by the analysis of the CpG methylation status of genes selected from the group consisting ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences.

15 In a further aspect, the present invention achieves said goal by analysis of the methylation status of at least one CpG position of the gene ALX4 and/or its regulatory sequences. In a further aspect the aim of the invention is achieved by the methylation analysis of said gene, ALX 4 and/or its regulatory sequences and one or more genes selected from the group consisting TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their  
20 regulatory sequences.

The present invention is further based upon the analysis of methylation levels within two or more genes taken from the group consisting of **ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences.**

Accordingly, the invention also disclose the genomic sequences of said genes in SEQ ID  
25 NO: 1 TO SEQ ID NO: 4 AND SEQ ID NO: 45 TO SEQ ID NO: 47, according to table 2 . Additional embodiments provide modified variants of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine methylation patterns within SEQ ID NO:1 to SEQ ID NO:4 and SEQ ID NO:45 to SEQ ID NO:47.

30 According to the present invention hypermethylation of the genes ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences is correlated with varying degrees of probability to the presence of colon cell proliferative disorders, and or metastases thereof. The present invention discloses the analysis of methylation within said genes and/or their regulatory sequences in the form of a panel enabling the improved detection,

classification, treatment and overall prognosis of colon cell proliferative disorders. Aberrant methylation of the genes TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 have to date been associated with the development of colorectal cell proliferative disorders. The present invention provides specific combinations of these genes which were determined to be particularly useful for the detection of colorectal cell proliferative disorders as measured by sensitivity and specificity of detection. Furthermore the invention provides CpG methylation analysis of the gene ALX4, with specific and novel utility for the detection of colorectal cell proliferative disorders. Methylation analysis of this gene is herein shown to have the surprising effect of being a highly sensitive and specific colorectal cancer detection marker. Furthermore the sensitivity and specificity of this detection is improved by a combined analysis of the gene ALX4 and one or more genes selected from the group consisting of TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3.

Wherein the object of the analysis is the detection of colon cell proliferative disorders it is particularly preferred that the methylation of two or more genes selected from the group consisting of ALX4, TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences are analysed. It is particularly preferred that said genes are selected from the group consisting of ALX4, TPEF and p16 and/or their regulatory sequences are analysed. It is further preferred that the methylation of all of the genes of the group consisting ALX4, TPEF and p16 and/or their regulatory sequences are analysed. In an alternative embodiment the methylation of the gene ALX4 only is analysed. In a further preferred alternative embodiment the CpG methylation status of the gene ALX4 and/or its regulatory sequences and one or more genes selected from the group consisting of TPEF, p16/INK4A, APC, caveolin-2, DAPK and TIMP3 and/or their regulatory sequences are analysed.

An objective of the invention comprises analysis of the methylation state of two or more CpG dinucleotides within at least two of the genomic sequences selected from the group consisting of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 and sequences complementary thereto.

It is preferred that the methylation of two or more sequences selected from the group consisting SEQ ID NOS: 2, 3 and 4 are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the sequences of the group consisting SEQ ID NOS: 2, 3 and 4 are analysed.

In an alternative embodiment the methylation status of at least one CpG position of SEQ ID NO: 2 only is analysed. In a further preferred alternative embodiment the CpG methylation status of SEQ ID NO:2 and one or more sequences selected from the group consisting SEQ ID

NOS:1, 3, 4 and SEQ ID NOS:45 to SEQ ID NO:47 and sequences complementary thereto are analysed. In a further preferred alternative embodiment the CpG methylation status of SEQ ID NO:2 and one or more sequences selected from the group consisting SEQ ID NOS:1, 3 and 4 and sequences complementary thereto are analysed.

5       The disclosed invention provides treated nucleic acids, derived from genomic SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization. The genomic sequences in question may comprise one, or more, consecutive or random methylated  
10   CpG positions. Said treatment preferably comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof. In a preferred embodiment of the invention, the objective comprises analysis of at least two modified nucleic acid comprising a sequence of at least 16 contiguous nucleotide bases in length of a sequence selected from the group consisting of SEQ ID NO: 5 TO SEQ ID NO: 20 & SEQ ID NO: 48 TO  
15   SEQ ID NO: 59, wherein said sequence comprises at least one CpG, TpA or CpA dinucleotide and sequences complementary thereto. The sequences of SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20 provide modified versions of the nucleic acid according to SEQ ID NOS:2 to SEQ ID NO:4, wherein the modification of each genomic sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from said genomic  
20   sequence as follows. For each sense strand genomic DNA, e.g., SEQ ID NO:1, four converted versions are disclosed. A first version wherein "C" is converted to "T," but "CpG" remains "CpG" (*i.e.*, corresponds to case where, for the genomic sequence, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); a second version discloses the complement of the disclosed genomic DNA sequence (*i.e.* *antisense* strand), wherein "C" is  
25   converted to "T," but "CpG" remains "CpG" (*i.e.*, corresponds to case where, for all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted). The 'upmethylated' converted sequences of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 correspond to SEQ ID NOS:5 to SEQ ID NO:12 and SEQ ID NOS:48 to SEQ ID NO:53 (see TABLE 2). A third chemically converted version of each genomic sequences is  
30   provided, wherein "C" is converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (*i.e.*, corresponds to case where, for the genomic sequences, all "C" residues of CpG dinucleotide sequences are unmethylated); a final chemically converted version of each sequence, discloses the complement of the disclosed genomic DNA sequence (*i.e.* *antisense* strand), wherein "C" is converted to "T" for all "C" residues, including those of



“CpG” dinucleotide sequences (*i.e.*, corresponds to case where, for the complement (*antisense* strand) of each genomic sequence, all “C” residues of CpG dinucleotide sequences are unmethyated). The ‘downmethyated’ converted sequences of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 correspond to SEQ ID NOS:13 to SEQ ID NO:20 and  
5 SEQ ID NOS:54 to SEQ ID NO:59.

Particularly useful for the detection of colon cell proliferative disorders, and heretofore undisclosed are the non-naturally occurring sequences according to SEQ ID NOS:7, 8, 15 and 16, which correspond to methylation-specific converted sequences of part of the gene ALX4 (SEQ ID NO:2).

10 In an alternative preferred embodiment, such analysis comprises the use of an oligonucleotide or oligomer for detecting the cytosine methylation state within genomic or pretreated (chemically modified) DNA, according to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:48 to SEQ ID NO:59. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridizes, under moderately  
15 stringent or stringent conditions (as defined herein above), to a pretreated nucleic acid sequence according to SEQ ID NOS:5 to SEQ ID NO:20 & SEQ ID NOS:48 to SEQ ID NO:59 and/or sequences complementary thereto, or to a genomic sequence according to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (*e.g.*, oligonucleotides and  
20 peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridize under moderately stringent and/or stringent hybridization conditions to all or a portion of the sequences SEQ ID NOS:1 to SEQ ID NO:20, or to the complements thereof. The hybridizing portion of the hybridizing nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present  
25 invention.

Preferably, the hybridizing portion of the inventive hybridizing nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NOS:1 to SEQ ID NO:20, or to the complements thereof.

Hybridizing nucleic acids of the type described herein can be used, for example, as a  
30 primer (*e.g.*, a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridization of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or  $T_m$ , which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required

stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest  
 5 temperature at which only homologous hybridization occurs with a particular concentration of salt (e.g., SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C decrease in the  $T_m$ , the temperature of the final wash in the hybridization reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash temperature is decreased by 5 °C). In practice, the change in  $T_m$  can be between 0.5°C and  
 10 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, e.g., SEQ ID NO:1, include those corresponding to sets (sense and antisense sets) of consecutively overlapping oligonucleotides of length X, where the oligonucleotides within each consecutively overlapping set (corresponding to a given X  
 15 value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

n to (n + (X-1));

where n=1, 2, 3,...(Y-(X-1));

where Y equals the length (nucleotides or base pairs);

where X equals the common length (in nucleotides) of each oligonucleotide in  
 20 the set (e.g., X=20 for a set of consecutively overlapping 20-mers); and where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to Y-(X-1).

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

25 Examples of inventive 20-mer oligonucleotides within a sequence of length 2470 base pairs include the following set of 2470 oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions 1-20, 2-21, 3-22, 4-23, 5-24, .....2451-2470.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

30 The present invention encompasses, for *each* of SEQ ID NO: 1 to SEQ ID NO: 20 and SEQ ID NO: 45 to SEQ ID NO: 59 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X, where, e.g., X= 9, 10, 17, 20, 22, 23, 25, 27, 30 or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective

tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO 47. Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO:1 to SEQ ID NO:20  
5 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where  
10 the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores,  
15 fluorophors, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide  
20 may include other appended groups such as peptides, and may include hybridization-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a chromophore, fluorophor, peptide, hybridization-triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

25 The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligonucleotides or oligomers according to particular embodiments of the present invention are typically used in 'sets,' which contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequence SEQ ID NO: 2 to SEQ ID NO: 4 and sequences  
30 complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the pretreated nucleic acids according to SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20 and sequences complementary thereto. In a further preferred embodiment the set comprises contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequence SEQ ID NO:2 and sequences complementary thereto,

or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the pretreated nucleic acids according to SEQ ID NOS: 7, 8, 15 and 16.

However, it is anticipated that for economic or other factors it may be preferable to analyze a limited selection of the CpG dinucleotides within said sequences, and the content of the set of oligonucleotides is altered accordingly.

Therefore, in particular embodiments, the present invention provides a set of at least two (2) (oligonucleotides and/or PNA-oligomers) useful for detecting the cytosine methylation state in pretreated genomic DNA of at least two genes selected from ALX4, TPEF and p16 (SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20), or in genomic DNA (SEQ ID NOS:2 to SEQ ID NO:4 and sequences complementary thereto). In a further preferred embodiment the set comprises at least two oligonucleotides for the analysis of CpG positions within one or more of SEQ ID NOS:7, 8, 15 and 16.

These probes enable diagnosis, classification and/or therapy of genetic and epigenetic parameters of colon cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in pretreated genomic DNA (SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20), or in genomic DNA (SEQ ID NOS:2 to SEQ ID NO:4 and sequences complementary thereto).

In preferred embodiments, at least one, and more preferably all members of a set of oligonucleotides is bound to a solid phase.

In further embodiments, the present invention provides a set of at least two (2) oligonucleotides that are used as 'primer' oligonucleotides for amplifying DNA sequences of one of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:49 and sequences complementary thereto, or segments thereof.

It is anticipated that the oligonucleotides may constitute all or part of an "array" or "DNA chip" (*i.e.*, an arrangement of different oligonucleotides and/or PNA-oligomers bound to a solid phase). Such an array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized, for example, in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid-phase surface may be composed of silicon, glass, polystyrene, aluminum, steel, iron, copper, nickel, silver, or gold. Nitrocellulose as well as plastics such as nylon, which can exist in the form of pellets or also as resin matrices, may also be used. An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of Nature Genetics (*Nature Genetics Supplement*, Volume 21, January 1999, and from the literature cited therein). Fluorescently labeled probes are often used for the scanning of immobilized DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the



specific probe are particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridized probes may be carried out, for example, via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

It is particularly preferred that the oligomers according to the invention are utilised for the detection of colorectal carcinoma.

The present invention further provides a method for ascertaining genetic and/or epigenetic parameters of the genes ALX4, TPEF, p16, APC, TIMP3, Dapk and Caveolin 2 and/or their regulatory sequences within a subject by analyzing cytosine methylation and single nucleotide polymorphisms.

It is preferred that the methylation of two or more genes selected from the group consisting of the genes ALX4, TPEF, p16, APC, TIMP3, Dapk and Caveolin 2 and/or their regulatory sequences are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the genes of the group consisting ALX4, TPEF and p16 and/or their regulatory sequences are analysed.

In a further embodiment of the method it is preferred that only the methylation status of the gene ALX4 and/or its regulatory sequences is analysed. In a further preferred embodiment the methylation status of the gene ALX4 and/or its regulatory sequences and one or more of the group consisting TPEF, p16, APC, TIMP3, Dapk and Caveolin 2 and/or their regulatory sequences are analysed. In a further preferred embodiment the methylation status of the gene ALX4 and/or its regulatory sequences and one or both of the group consisting TPEF and p16 and/or their regulatory sequences are analysed.

Accordingly, it is preferred that the methylation of two or more genetic sequences selected from the group consisting SEQ ID NOS: 2,3 and 4 and/or their regulatory sequences are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the genes of the group consisting SEQ ID NOS: 2,3 & 4 and/or their regulatory sequences are analysed.

In a further embodiment of the method it is preferred that only the methylation status of SEQ ID 2 is analysed. In a further preferred embodiment the methylation status of SEQ ID 2 and one or more of the group consisting SEQ ID NOS: 3, 4, 45 – 47 are analysed. In a further preferred embodiment the methylation status of SEQ ID NO:2 and one or both of SEQ ID NOS:3 and 4 are analysed.

Said method comprising contacting a nucleic acid comprising one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 in a biological sample obtained

from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

It is preferred that the methylation of two or more genetic sequences select analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the genes of the group consisting SEQ ID NOS:2, 3 and 4 and/or their regulatory sequences are analysed.

In a further embodiment of the method it is preferred that only the methylation status of SEQ ID 2 is analysed. In a further preferred embodiment the methylation status of SEQ ID NO:2 and one or more of the group consisting SEQ ID NOS:3, 4 45 – 47 are analysed. In a further preferred embodiment the methylation status of SEQ ID NO:2 and one or both of SEQ ID NOS:3 and 4 are analysed.

Said method comprising contacting a nucleic acid comprising the appropriate gene(s) and/or one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 in a biological sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

Preferably, said method comprises the following steps: In the *first step*, a sample of the tissue to be analysed is obtained. The source may be any suitable source, such as cell lines, histological slides, biopsies, tissue embedded in paraffin, bodily fluids, stool, blood and all possible combinations thereof. Genomic DNA is then isolated from said biological sample, this may be by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants, *e.g.*, by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense and required quantity of DNA.

Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

In the *second step* of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be

understood as 'pretreatment' herein.

The above described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behavior.

In the *third step* of the method, fragments of the pretreated DNA are amplified, using sets of primer oligonucleotides according to the present invention, and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO:20 and sequences complementary thereto.

It is preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NOS:7-12 and 15-20. In this embodiment of the invention it is particularly preferred that said set consists of at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of SEQ ID NOS:5-20 and 48-59.

It is also preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NOS:7-12 and 15-20.

It is also preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of at least one of SEQ ID NOS:7, 8, 15 or 16. It is also preferred that said set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of at least one of SEQ ID NOS:7, 8, 15 or 16, and a further pair of oligonucleotides whose sequences are each reverse complementary,

identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of at least one of SEQ ID NOS:9-12 and 17-20.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

It is preferred that said set of MSP primer oligonucleotides includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7-12, 15-20 and 49-59.

In this embodiment of the invention it is particularly preferred that said set consists of at least two MSP primer oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of SEQ ID NOS: 7-12 and 15-20.

In a further embodiment of the invention it is particularly preferred that said set consists of at least two MSP primer oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of at least one of the base sequences of SEQ ID NOS:7, 8, 15 and 16. In a further preferred embodiment of this method it is preferred that said set consists of at least two MSP primer oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of at least one of the base sequences of SEQ ID NOS:7, 8, 15 and 16 and one or more sequences taken from the group consisting SEQ ID NOS:9-12, 17-20 and 48-59.

A further preferred embodiment of the method comprises the use of *blocker*



oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. Blocking probe oligonucleotides are hybridized to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a  
5 nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking  
10 probes comprising a 'CpA' or 'TpA' at the position in question, as opposed to a 'CpG' if the suppression of amplification of methylated nucleic acids is desired.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-  
15 deoxyoligonucleotides, or oligonucleotides derivitized at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate  
20 bridges at the 5'-termini thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (e.g., with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is  
25 because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker—a process that normally results in degradation of the hybridized blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as  
30 blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

Preferably, therefore, the base sequence of said *blocking oligonucleotides* is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID

NOS:48 to SEQ ID NO:59 and sequences complementary thereto, wherein the base sequence of said oligonucleotides comprises at least one CpG, TpG or CpA dinucleotide.

It is preferred that said set of *blocking oligonucleotides* includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides  
5 which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7-12 and 15-20. In this embodiment of the invention it is particularly preferred that said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7-12 and 15-20.

10 In a further embodiment it is preferred that said set of *blocking oligonucleotides* includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7, 8, 15 & 16. In this embodiment of the invention it is particularly preferred that said set further consists of at least one blocker oligonucleotides whose sequences are each  
15 reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:9-12, 17-20 & 48-59. In this embodiment it is further preferred that said base sequences are selected from SEQ ID NOS:9-12 and 17-20.

The fragments obtained by means of the amplification can carry a directly or indirectly  
20 detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplicates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of, e.g., matrix assisted  
25 laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas & Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is  
30 evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The

analysis of nucleic acids is somewhat more difficult (Gut & Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionization process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut & Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In the *fourth step* of the method, the amplicates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplicates were obtained by means of MSP amplification, the presence or absence of an amplicate is in itself indicative of the methylation state of the CpG positions covered by the primer, according to the base sequences of said primer.

Amplicates obtained by means of both standard and methylation specific PCR may be further analyzed by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplicates synthesised in *step three* are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplicates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse

complementary or identical to a segment of the base sequences specified in the present Sequence Listing; and the segment comprises at least one CpG, TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said  
5 dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, and the equivalent positions within SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO: 20. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-  
10 hybridized amplicates are then removed. The hybridized amplicates are then detected. In this context, it is preferred that labels attached to the amplicates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridised to the  
15 bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; also see United States Patent No. 6,331,393) employing a dual-labeled fluorescent oligonucleotide probe (TaqMan™  
20 PCR, using an ABI Prism 7700 Sequence Detection System, Perkin Elmer Applied Biosystems, Foster City, California). The TaqMan™ PCR reaction employs the use of a nonextendible interrogating oligonucleotide, called a TaqMan™ probe, which, in preferred imbodiments, is designed to hybridize to a GpC-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety"  
25 and a "quencher moiety" covalently bound to linker moieties (e.g., phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLight™ assay. Variations on the TaqMan™ detection  
30 methodology that are also suitable for use with the described invention include the use of dual-probe technology (Lightcycler™) or fluorescent amplification primers (Sunrise™ technology). Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

A further suitable method for the use of probe oligonucleotides for the assessment of



methylation by analysis of bisulfite treated nucleic acids In a further preferred embodiment of the method, the *fifth step* of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

5 In yet a further embodiment of the method, the *fifth step* of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the *third step* of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

#### Best mode

10 In the most preferred embodiment of the method the nucleic acids according to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NO: 45 to SEQ ID NO: 47 are isolated and treated according to the first three steps of the method outlined above, namely:

- a. obtaining, from a subject, a biological sample having subject genomic DNA;
- b. extracting or otherwise isolating the genomic DNA;
- 15 c. treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;

and wherein the subsequent amplification of d) is carried out in a methylation specific manner, namely by use of methylation specific primers or *blocking oligonucleotides*, and further wherein  
20 the detection of the amplificates is carried out by means of a real-time detection probes, as described above.

It is particularly preferred that the methylation of one or more sequences selected from the group consisting SEQ ID NOS: 2,3 and 4 are analysed. In this embodiment of the invention  
25 it is further preferred that the methylation of all of the sequences of the group consisting SEQ ID NOS: 2,3 and 4 are analysed.

In a further embodiment it is preferred that only the methylation status of SEQ ID NO:2 is analysed. In a further embodiment it is preferred that the methylation status of SEQ ID NO:2 and one or more sequences selected from the group consisting SEQ ID NOS:3, 4, 45-47 are  
30 analysed, even more preferably said group consists SEQ ID NOS:3 and 4 only.

Wherein the subsequent amplification of d) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NOS:5 to SEQ ID NO:20 and SEQ ID NOS:48 to SEQ ID NO:59

and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

It is preferred that said set of MSP primer oligonucleotides includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides  
5 which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7-12 and 15-20.

In a further embodiment it is preferred that the said set MSP primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair  
10 long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 & 16.

In a further preferred variant of said embodiment said set consists of at least MSP primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 and 16  
15 and at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:9-12 and 17 – 20.

Step e) of the method, namely the detection of the specific amplicates indicative of the methylation status of one or more CpG positions according to SEQ ID NOS:1 to SEQ ID NO:4  
20 and SEQ ID NOS:45 to SEQ ID NO:47 is carried out by means of real-time detection methods as described above.

In an alternative most preferred embodiment of the method the subsequent amplification of d) is carried out in the presence of *blocking oligonucleotides*, as described above. Said *blocking oligonucleotides* comprising a sequence having a length of at least 9 nucleotides which  
25 hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NOS:5-20 and 48-59 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide.

It is preferred that said set of *blocking oligonucleotides* includes at least two oligonucleotides whose sequences comprise a sequence having a length of at least 9 nucleotides  
30 which hybridizes to a pretreated nucleic acid sequence according to one or more of SEQ ID NOS:7 to SEQ ID NO:12 and SEQ ID NOS:15 to SEQ ID NO: 20. In this embodiment of the invention it is particularly preferred that said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each

of SEQ ID NOS:7-12 and 15-20.

In a further embodiment it is preferred that the said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 and 16.

In a further preferred variant of said embodiment said set consists of at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:7, 8, 15 and 16 and at least one blocker oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 9-base-pair long segment of the base sequences of each of SEQ ID NOS:9-12 and 17-20.

Step e) of the method, namely the detection of the specific amplicates indicative of the methylation status of one or more CpG positions according to SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 is carried out by means of real-time detection methods as described above.

Additional embodiments of the invention provide a method for the analysis of the methylation status of genomic DNA according to the invention (SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, and complements thereof) without the need for pretreatment.

It is preferred that the methylation of one or more sequences selected from the group consisting SEQ ID NOS: 2, 3 and 4 are analysed. In this embodiment of the invention it is particularly preferred that the methylation of all of the sequences of the group consisting SEQ ID NOS: 2, 3 and 4 are analysed.

In the *first step* of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, body fluids, stool or tissue embedded in paraffin. In the *second step*, the genomic DNA is extracted. This may be by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants e.g. by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time,

expense and required quantity of DNA.

In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, in particular with methylation-sensitive restriction endonucleases.

5 In the *third step*, the DNA is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide.

In the *fourth step*, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and  
10 said amplicates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels.

In the *fifth step* the amplicates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridization analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI  
15 or ESI analysis.

In the final step of the method the presence or absence of colon cell proliferative disorder is deduced based upon the methylation state of at least one CpG dinucleotide sequence of SEQ ID NOS:1 to SEQ ID NO:4 & SEQ ID NOS:45 to SEQ ID NO:47, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of  
20 SEQ ID NOS:1 to SEQ ID NO 4 and SEQ ID NOS:45 to SEQ ID NO:47.

In a further embodiment said deduction is based upon the methylation status of SEQ ID NO:2 only. In a further preferred embodiment said deduction is based upon the methylation status of SEQ ID 2 and one or more sequences chosen from, SEQ ID NOS:3 and 4.

## 25 Diagnostic Assays for colon cell proliferative disorders

The present invention enables diagnosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within one or more of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 may be used as markers. Said parameters obtained by means of the present invention may be compared to  
30 another set of genetic and/or epigenetic parameters, the differences serving as the basis for a diagnosis and/or prognosis of events which are disadvantageous to patients or individuals.

Specifically, the present invention provides for diagnostic cancer assays based on measurement of differential methylation of one or more CpG dinucleotide sequences of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47, or of subregions thereof that



comprise such a CpG dinucleotide sequence. Typically, such assays involve obtaining a tissue sample from a test tissue, performing an assay to measure the methylation status of at least one of one or more CpG dinucleotide sequences of SEQ ID NOS:1 to SEQ ID NO:4 and SEQ ID NOS:45 to SEQ ID NO:47 derived from the tissue sample, relative to a control sample, or a  
5 known standard and making a diagnosis or prognosis based thereon.

In particular preferred embodiments, inventive oligomers are used to assess the CpG dinucleotide methylation status, such as those based on SEQ ID NOS:1 to SEQ ID NOS: 20 and 45 to 59, or arrays thereof, as well as in kits based thereon and useful for the diagnosis and/or prognosis of colon cell proliferative disorders.

10

#### Kits

Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent; a set of primer oligonucleotides containing at least two oligonucleotides whose sequences in each case correspond, are complementary, or hybridize  
15 under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NOS:1 to SEQ ID NO:20 & SEQ ID NOS:45 to SEQ ID NO:59; oligonucleotides and/or PNA-oligomers; as well as instructions for carrying out and evaluating the described method.

More preferred is a kit comprising the oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a  
20 16-base long segment of the sequences SEQ ID NOS: 5-20.

Also preferred is a kit comprising the oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NOS: 7, 8, 15 & 16.

Further preferred is a kit comprising the oligonucleotides whose sequences in each case  
25 correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of the sequences SEQ ID NOS:7, 8, 15 and 16 and at least one of SEQ ID NOS:9-12 and 17-20.

In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one  
30 or more of the following techniques: MS-SNuPE<sup>TM</sup>, MSP, MethyLight®, HeavyMethyl<sup>TM</sup>, COBRA<sup>TM</sup>, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

Typical reagents (e.g., as might be found in a typical COBRA<sup>TM</sup>-based kit) for COBRA<sup>TM</sup> analysis may include, but are not limited to: PCR primers for specific gene (or

methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (e.g., as might be found in a typical MethyLight®-based kit) for MethyLight® analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

Typical reagents (e.g., as might be found in a typical Ms-SNuPE™-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE™ primers for specific gene; reaction buffer (for the Ms-SNuPE™ reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (e.g., as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following example serves only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

## EXAMPLES

### Material and Methods

#### Subjects for MethyLight Analysis

Colon tissues were obtained by surgical resection from 47 patients (29 male, 18 female) with colon cancer, with a median age of 66 years (range 31-93 years), from the tumor and a tumor-free location which was at least 2 cm distant from the tumor and which was confirmed to be without any tumor cell infiltration by histological assessment. In all 47 patients tissue samples from the colon cancer were obtained for molecular analysis, in 21 of these cases a

matched non-cancer colon sample was also obtained for molecular analysis after tumor cell infiltration was ruled out by histological assessment. The metastatic lesions were obtained from 24 patients (13 male, 11 female, median age 64.5 yrs, range 41-79) with colorectal cancer that developed liver metastasis after prior successful colon cancer resection. In one case the primary colon cancer and a single liver metastasis were resected at the same time in a 74 year old female patient. Immediately after surgery, tissue samples were put in liquid nitrogen and stored at -80 °C until use. Formalin fixed tissues were processed as previously described and sections were stained with hematoxylin and eosin for histological evaluation. Tumor stages were assessed using the TNM-system.

#### DNA Extraction

Genomic DNA was extracted from the tissues using the proteinase K digestion method.

#### Genome-Wide Methylation Screening Assay

Differentially methylated genomic sequences were identified using Methylation-specific arbitrarily primed PCR analysis (MS AP-PCR) by comparison of different levels of disease to age-matched normal tissue for several different age groups (see for example Young J, Biden KG, Simms LA, Huggard P, Karamatic R, Eyre HJ, Sutherland GR, Herath N, Barker M, Anderson GJ, Fitzpatrick DR, Ramm GA, Jass JR, Leggett BA. HPP1: a transmembrane protein-encoding gene commonly methylated in colorectal polyps and cancers. Proc Natl Acad Sci USA 2001;98:265-270.) DNA from colon adenomas, CRC samples from patients with no lymph node involvement or metastasis and CRC samples from patient with lymph node involvement and/or metastasis was compared in MS AP-PCR experiments to age-matched normal colon tissue and to age-matched normal peripheral blood lymphocytes.

Five samples for each tissue type were collected for each of three age groups; over 65 years, 50 to 65 years and under 50 years. Genomic DNA was extracted from the tissue samples using Qiagen Genomic-Tip 500/G columns. The five DNA samples from each tissue type and each age group were pooled and experiments were performed as follows. DNA was digested with *RsaI* to generate smaller DNA fragments before digestion with *MspI* and *HpaII*, two restriction enzymes with different sensitivities to cytosine methylation. Then MS AP-PCR was performed as previously described by Liang et al. The digested DNA was amplified using these sets of primers: G1, 5'-GCGCCGACGT-3'; G5, 5'-TGCGACGCCG-3'; APBS5, 5'-CTCCCACGCG-3'. After amplification fragments were separated on polyacrylamide gels and those exhibiting a pattern of differential methylation were eluted from the gel, cloned into

vectors and sequenced as outlined above. Identification of sequences was performed by BLAST searches in Genbank.

### MethyLight™ Analysis

5 Genomic DNA was analyzed by the MethyLight technique after bisulfite conversion as previously reported by Eads et al. (Epigenetic patterns in the progression of esophageal adenocarcinoma. *Cancer Res* 2001;61:3410-3418. and Fields of aberrant CpG island hypermethylation in Barrett's esophagus and associated adenocarcinoma. *Cancer Res* 2000;60:5021-5026.). In this analysis three oligos are used in every reaction. Two locus-specific PCR primers flank an oligonucleotide probe with a 5' fluorescent reporter dye (6FAM) and a 3' quencher dye (BHQ-1). For this analysis primers and probes are specifically designed to bind to bisulfite-converted DNA, which generally span 7 to 10 CpG dinucleotides. The gene of interest is then amplified and normalized to a reference set ( $\beta$ -actin (*ACTB*)) to normalize for input DNA. The specificity of the reactions for methylated DNA is confirmed using human sperm DNA (unmethylated) and CpGenome Universal Methylated DNA (Chemicon (subsidiary of Serologicals) catalog #S7821) (methylated). For standardization the primers and the probe for analysis of the *ACTB* gene lack CpG dinucleotides so that amplification is possible regardless of methylation levels. TaqMan PCR reactions were performed in parallel with primers specific for the bisulfite-converted methylated sequence for a particular locus and with the *ACTB* reference primers. The ratio between the values was calculated in these two TaqMan analyses, using this approach the degree of methylation at that locus was determined. The extent of methylation at a specific locus was determined by the following formula:

$$[(\text{gene/actb})^{\text{sample}} : (\text{gene/actb})^{\text{SssI-treated genomic DNA}}] \times 100.$$

A cut off value of 4% gave the best discrimination between normal and cancerous samples, as previously reported. Therefore, samples with  $\geq 4\%$  fully methylated molecules were termed methylated, whereas samples with  $< 4\%$  were considered unmethylated. The primer and probe sequences are listed in Table 1 and were used as previously reported by Eads et al.

### Bisulfite Sequencing

30 Bisulfite genomic sequencing was performed for the ALX4 gene in order to confirm the results obtained by MethyLight™ analysis. Briefly, bisulfite treated genomic DNA from 4 colon cancers and matched normal colon mucosa was amplified with primers specific for a



fragment of the ALX4 gene containing 39 CpG sites and spanning the region that was analysed with the MethyLight<sup>TM</sup> assay: ALX4\_bis1, 5'-TGAATAGGGTGATATTTTAGTTAGG-3'; ALX4\_bis2, 5'-ATAAATCATCCCAAACCTCTA-3' (SEQ ID NO: 60). PCR was carried out in a reaction mixture (25 µl) containing 7 µl of DNA, 0.2 mM dNTPs, 1 µM primers, and 0.25 units of DyNAzyme EXT DNA Polymerase (Finnzymes). Amplification was performed using the following condition: 94 °C for 2 min, followed by 36 cycles (94 °C for 1 min, 45 °C for 1 min, 72 °C for 2 min) and then 72 °C for 10 minutes. PCR products were separated on 1% agarose gel, stained with ethidium bromide and visualized with an UV transilluminator. DNA fragments of interest were cloned into a plasmid vector with the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) according to manufacturer's recommendations and sequence was confirmed by automated sequencing.

### Statistical Analysis

The PMR values of the MethyLight assays were dichotomized for statistical purposes as previously reported by Eads et al.. PMR values above 4% were considered as methylation positive and classified as '1', whereas PMR levels below 4% were classified as '0' (no methylation). This dichotomization should level off the quantitative impact of different levels of hypermethylation per gene, and allow the cross-gene comparison of methylation per gene in colon cancer and metastasis. The different clinicopathological features, such as location of primary tumor, grade of differentiation or stage of cancer were used as nominal variables in the Fisher's exact test or Chi square test. Otherwise student's t-test was used to determine statistical difference. All tests were two-sided, and a *p*-value of <0.05 was considered statistically significant.

### Results

Differential methylation of a fragment of approximately 242 bp with genomic sequence matching a portion of the first intron of ALX4 was confirmed in seven different MS AP-PCR experiments comparing colon cancer and adenoma DNA to normal DNA obtained from the same individuals. In experiments performed on samples from patients over 65 years old, bands corresponding to a methylated fragment of ALX4 were found in DNA from adenomas, non-metastatic adenocarcinoma and metastatic adenocarcinoma samples when compared to normal age-matched colon tissue. The band was also identified in a mixture of pre-cancerous and adenocarcinoma samples compared to age-matched PBL DNA. An identical band was found in adenoma sample DNA from patients 50-65 years old when compared to normal age-matched

colon tissue and a mixture of pre-cancerous and adenocarcinoma DNA from patients 50-65 years compared to age-matched PBL DNA. Lastly the fragment was found in a comparison of adenomas from patients less than 50 years compared to a ge-matched normal tissue (data not shown).

5

#### Confirmation of ALX4 gene methylation in primary and metastatic colorectal cancer

The methylation of the ALX4 gene was assessed in 47 colon cancers and compared to 21 cases of normal colon mucosa which were obtained from a subset of these cancer patients. Using the MethyLight assay a high degree of methylation was found in the cancerous colon as compared to the matched normal colon mucosa (Figure 2A). Thirty cancers exhibited a PMR > 4% (30/47) whereas in none of the 21 normal colon samples ALX4 gene methylation was observed ( $p < 0.0001$ ). We then assessed the degree of ALX4 methylation in our series of metastatic samples and compared them to the primary colon cancers. In this series, apart from one patient, the tissues were, however, not matched and thus from different patients. Nonetheless, our analysis showed that a high degree of methylation can be found in both primary (30/47) and metastatic colon cancer (16/24), which did not differ by statistical analysis (Figure 2B). Interestingly, primary colon cancer and metastatic tissues from the one female patient from which both tissues were obtained, exhibited a low degree of methylation in the primary cancer and a high degree of methylation in the metastatic lesion (Figure 2B).

In order to confirm the results of ALX4 gene methylation obtained by MethyLight assay we also performed bisulfite sequencing on 4 matched normal and colon cancer samples. These 4 patients were selected because there was a dramatic difference in the levels of methylation of the ALX4 gene in the cancer versus the matched normal colon sample (Figure 3). The region of the ALX4 gene that was analysed by bisulfite sequencing spanned 39 CpG sites, including the 12 CpG sites that were analysed with the MethyLight assay. The DNA fragments encoding ALX4 were amplified by PCR using bisulfite treated DNA and 3-11 clones per sample were sequenced. Similar to the results obtained with the MethyLight assay we found that the majority of CpG sites were methylated in the cancer samples, whereas in the normal colon mucosa the CpG sites were widely unmethylated (Figure 3). The varying degrees of methylation that were observed, however, in the cancer samples reflect the degree of contaminating non-malignant cells that were present in the DNA preparation from the cancer tissues used for methylation analysis. Overall, the results obtained by bisulfite sequencing confirmed the results of the PCR based assay, indicating that the MethyLight assay correctly assesses the methylation of the ALX4 gene in colorectal cancer.



### Analysis of gene methylation in primary and metastatic colorectal cancer

Next we analysed our tissues for gene methylation using a set of genes that had been previously be reported by other groups to be associated with either colon cancer pathogenesis or the development of cancer metastasis: TPEF, p16/INK4A, APC, caveolin-2, DAPK, TIMP3. Using MethyLight assays we assessed the methylation status of ALX4 and the other 6 genes in our series of primary and metastatic colon cancers. The cut-off of methylation was chosen to be a PMR (percentage of methylated reference) of >4% (as previously reported by Eads et al.) and all samples with a PMR >4% were classified as methylation positive ('1'), whereas samples with a PMR below 4% were considered methylation negative ('0'). The results of the methylation analysis of each gene in primary and colon cancer is given in table 3. In addition, the numbers were added giving the total numbers of methylated genes per sample. Using this approach we observed at least one methylated gene in 40 of 47 primary cancers, this would indicate a sensitivity of 85.1% for the detection of cancer (Figure 4A). In contrast, only one of 21 normal colon mucosa samples exhibited TPEF gene methylation, indicating that the specificity of this marker set would be 95.2%. Thus, the analysis of only three genes, *i.e.* ALX4, TPEF and p16, allowed the identification of these 40 cancers, whereas the other genes did not further contribute to the detection of colon cancer. In the metastatic lesions all 24 cases exhibited at least one of these methylated gene: ALX4, TPEF or APC, accounting for a detection rate of 100% (Figure 4B).

20

### ALX4 contributes to the identification of CIMP in primary and metastatic colon cancer

A condition termed CpG island methylator phenotype (CIMP) which is associated with microsatellite instability related to hMLH1 methylation, a proximal location in the colon and a family history of colon cancers was recently identified by Toyota et al. (CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA 1999;96:8681-8686.). In our series CIMP+ cancers were observed in 20 of 47 cases (Figure 4A). These cancers exhibited at least 3 methylated genes, which included ALX4, TPEF, APC, Caveolin-2, TIMP3 and p16. In the metastatic lesions we observed CIMP+ cancers in 9 of 24 cancers. The genes found to be methylated in CIMP+ metastatic lesions included ALX4, TPEF, TIMP3 and APC (Figure 4B). Colon cancers located in either the cecum, ascending colon, colon transversum or descending colon exhibited a CIMP+ status in 14 of 24 cases, whereas cancers of the sigma and/or rectum were CIMP+ in 6 of 22 cases, a difference that was also statistically significant ( $p=0.026$ ). However, CIMP+ tumors were not associated with either early or advanced stages in our series, and showed no association with the degree of differentiation.

30



Association of gene methylation in colon cancer with clinicopathological features of colon cancer

In order to assess a potential association of the presence of methylation with the location  
 5 of the primary tumor we classified our colorectal cancers into two groups: rectal cancers (n=10),  
 and non-rectal cancers (n=36). Using Fisher's exact test, we found that there was a statistical  
 significant difference in the presence of methylation with regard to the location of the primary  
 tumor. While rectal cancers exhibited no methylation in 4 of the cases, whereas methylation of  
 1 to 7 genes was found in 6 cases, the vast majority of colon cancers (34/36) exhibited  
 10 methylation in at least one gene ( $p=0.014$ ). Thus, from this analysis we can assume that  
 methylation is significantly more frequent in proximal, i.e. non-rectal cancers of the large  
 intestine (Figure 5A).

We analysed not only the association between the location of the primary and the overall  
 presence of gene methylation per patient, but also analysed each single gene with regard to this  
 15 association of location and gene methylation. However, apart from TPEF, none of the other  
 genes - including ALX4 - were linked to a certain location of the primary tumor, probably  
 because of the low number of methylated genes observed in our study. However, TPEF was  
 more frequently methylated in colon cancers (31/36) compared to rectal cancers (5/10)  
 ( $p=0.023$ ) (Figure 5B). The total number of methylated genes in rectal cancers (19/70) was also  
 20 compared to the total number of methylated genes in colon cancer (89/252), however, this  
 difference did not reach statistical significance ( $p=0.0513$ ).

While methylation is considered an early step in the process of colorectal cancer  
 pathogenesis, we assessed the presence of methylation in the cancers with regard to the stage of  
 the cancers. While none of our patients was in the UICC stage I, 20 patients presented with  
 25 UICC stage II cancer, 13 with stage III and 13 with stage IV colorectal cancer. Interestingly, the  
 number of patients without gene methylation increased with the UICC stage, in that patients  
 with UICC stage II had at least 1 methylated gene compared to the more advanced stages,  
 however this observation was not statistically significant. However, while we found no  
 association between the T stage and the presence of methylation, the presence of distant  
 30 metastasis (M+ stage) was associated with significantly less methylated cancers compared to  
 tumors without distant metastasis. Thus, while 3 of 6 cancers with distant metastasis did not  
 show any gene methylation, cancers without distant metastasis (M0) presented with at least one  
 methylated gene in 30 of 33 cases ( $p=0.033$ ) (Figure 5C). Again, the detailed analysis of every  
 single gene with regard to the association with distant metastasis did not identify a single gene

that was associated with the presence of distant metastasis, again maybe due to the limited number of cases analysed.

### Discussion

5 ALX4 gene methylation was identified by MS AP-PCR in our study and we confirmed the presence of ALX4 methylation in a larger series of primary and metastatic colon cancers using the MethyLight assay and bisulfite sequencing. ALX4 gene methylation was observed in 30 of 47 primary cancers and in none of the normal colon mucosa tissue samples. Furthermore, ALX4 was frequently methylated in the liver metastasis of a second set of patients with  
10 colorectal cancers. In the one female patient in which primary colon cancer and liver metastasis were resected at the same time we found a high degree of ALX4 gene methylation in the liver metastasis as opposed to the primary cancer, indicating that methylation of ALX4 may occur 'de novo' in the metastatic cancer cells even in the absence of methylation in the primary cancer cells.

15 In order to further assess the role of ALX4 gene methylation in primary and metastatic colon cancer, we analysed the presence of methylation of 6 other genes that have previously been linked either to colon cancer pathogenesis or metastatic development in our series of colon cancer and metastatic cancer tissues: TPEF/HPP1, p16/INK4A, APC, caveolin-2, DAPK and TIMP3. While TPEF and ALX4 exhibited a similar high frequency of methylation in primary  
20 colon cancer, p16 and APC were less frequently methylated. Of the 47 analysed primary colon cancers 40 exhibited at least one of the three methylated genes ALX4, TPEF or p16, indicating that a methylation based diagnostic test including these three genes may achieve a detection rate of 85.1%; since only one normal case exhibited a significant degree of methylation of the TPEF gene specificity would be 95.2%.

25 The cancers were then grouped according to the location of the primary cancer into rectal and non-rectal cancers. Overall gene methylation was more frequently present in non-rectal cancers compared to cancers of the rectum, an observation that has been reported by other groups as well. However, ALX4 gene methylation alone - in contrast to TPEF - was not associated with either location of the primary tumor, most probably due to the limited number of  
30 cases analysed in our study. Furthermore, since gene methylation has been considered to be an early event in the pathogenesis of this and other cancers, we assessed the frequency of methylation in each UICC stage and in cancers grouped according to the size of the primary tumor (T), presence of lymph nodes (N) or distant metastasis (M). However, colorectal cancers without distant metastasis frequently exhibited gene methylation, which contrasts with previous

reports that gene hypermethylation may be associated with a poor prognosis in this and other cancers.

Recently Toyota et al. (CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA 1999;96:8681-8686.) reported the presence of multiple methylated genes in colon cancer, a condition they termed CpG island methylator phenotype (CIMP). In their analysis two types of methylation were identified: Type A methylation referred to the age-related methylation of CpG islands in tumors and the normal colon mucosa. In contrast, some genes exhibited methylation only in colon cancers which was classified as type C methylation. Interestingly, most CIMP+ cancers were associated with microsatellite instability related to hMLH1 methylation, a proximal location and a family history of colon cancers. In our series CIMP+ cancers were observed in 20 of 47 cases and were located primarily in the proximal colon cancer, an association which has already been reported by Toyota et al. and Rijnsoever et al. (CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA 1999;96:8681-8686. and Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. Gut 2002;51:797-802.). However, our CIMP+ tumors were not associated with either early or advanced stages in our series, which is similar to the findings reported by Toyota et al.. In contrast to our study, Rijnsoever et al. reported that in their analysis CIMP+ colon cancers were poorly differentiated, a finding that was not observed in our and other studies and may be due to the larger patient group that was analysed in their study.

To our knowledge our study is the first to also address the role of gene methylation in metastatic colorectal cancers using a panel of 7 genes - including ALX4 - that were analysed by the highly sensitive MethyLight<sup>TM</sup> assay. Using this assay we found several genes to be methylated in both metastatic lesions and primary cancers, as well as genes that were neither methylated in metastasis nor in primary colorectal cancers. Apart from this observation, APC gene methylation increased in metastatic lesions compared to primary cancers. Based on these findings, we can classify the patterns of methylation in liver metastasis in three groups: class I genes: high degree of methylation in primary tumor and liver metastasis (ALX4, TPEF, p16), class II genes: higher degree of methylation in metastasis compared to primary tumor (APC) and class genes III: no methylation in either primary tumor or metastasis (caveolin-2, DAPK, TIMP3). Interestingly, all 24 metastases exhibited at least one methylated gene of the class I genes, indicating that this set of genes may be valuable for the methylation specific detection of liver metastasis in colon cancer. Interestingly, the two genes that exhibited a very high degree of methylation in the primary colon cancers, are also frequently methylated in the liver metastasis of colon cancer (ALX4, TPEF). From our analysis we assume that methylation of these genes

occurs early in these cancers and that they remain methylated in the progression of the disease. However, the high frequency of methylation detected in the primary cancers and metastatic lesions makes them ideal candidates for a methylation-based diagnostic tool for localized and metastatic colorectal cancer.

5 In summary, using MS AP-PCR we identified the methylation of ALX4 in colorectal cancers and further analysis revealed that ALX4 gene methylation is a frequent event in colorectal cancer pathogenesis. Together with a further set of genes, ALX4 allows the identification of primary and metastatic colorectal cancers indicating that methylation based diagnostic tests may be helpful in the identification of this and other malignancies and, thus,  
10 may improve the detection and overall prognosis of patients with these cancers.

Table 1. List of primers and probes used for MethyLight analysis

Gene	forward primer (5'-3')	reverse primer (5'-3')	probe sequence (5'-3')
ALX4	CGCGGTTTCGATTTTAATGC	ACTCCGACTTAACCCGACGAT	6FAM- CGACGAAATTCCTAACGCAACCGCTT AA-BHQ1
Caveol in 2	TTTCGGATGGGAACGGTGTA	CTCCACCGCCGTTACC	6FAM- CCCGTCCTAACCGTCCGCCCT-BHQ1
DAPK	TCGTCGTCGTTTCGGTTAGTT	CCCTCCGAAACGCTATCGA	6FAM- CGACCATAAACGCCAACGCCG-BHQ1
TPEF	TTTTTTTTTTCGGACGTCGTTG	CCCTCTACATACGCCGCGAAT	6FAM- AATTACCGAAAACATCGACCGA- BHQ1
p16/IN K4A	TGGAATTTTCGGTTGATTGGTT	AACAACGTCCGCACCTCCT	6FAM-ACCCGACCCCGAACCGCG- BHQ1
APC	GAACCAAACGCTCCCCAT	TTATATGTCGGTTACGTGCGTTTAT AT	6FAM- CCCGTCGAAAACCCGCCGATTA- BHQ1
TIMP3	GCGTCGGAGGTTAAGGTTGTT	CTCTCCAAAATTACCGTACGCG	6FAM-AACTCGCTCGCCCGCCGAA- BHQ1
Caveol in	TTTCGGATGGGAACGGTGTA	CTCCACCGCCGTTACC	6FAM- CCCGTCCTAACCGTCCGCCCT-BHQ1

15

Table 2. Genes and sequences according to the invention.



Gene name	Genomic SEQ ID NO	Methylated treated SEQ ID NOS:	Unmethylated treated SEQ ID NOS:
APC	1	5 & 6	13 & 14
ALX4	2	7 & 8	15 & 16
TPEF	3	9 & 10	17 & 18
p16	4	11 & 12	19 & 20
DAPK	45	48 & 49	54 & 55
TIMP3	46	50 & 51	56 & 57
Caveolin 2	47	52 & 53	58 & 59

Table 3. Summary of results from analysis of gene methylation in primary cancer and metastasis.

5

Gene	Normal (n=21)	Tumor (n=47)	Metastasis (n=24)	Class
ALX4	0/21	30/47*	16/24	I
TPEF	1/21	36/47*	19/24	I
p16	0/21	15/47 <sup>#</sup>	6/24	I
APC	0/21	10/47 <sup>#</sup>	10/24 <sup>s</sup>	II
TIMP3	1/21	11/47	2/24	III
DAPK	0/21	1/47	0/24	III
Caveolin 2	0/21	5/47	1/24	III